BEST AVAILABLE COPY

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
G06F 13/00, 12/00, 12/08
G06F 12/16, 11/00

A1

(11) International Publication Number: WO 94/00816

(43) International Publication Date: 6 January 1994 (06.01.94)

(21) International Application Number: PCT/US93/05853

(22) International Filing Date: 17 June 1993 (17.06.93)

(30) Priority data:
07/900,636 18 June 1992 (18.06.92) US

(71) Applicant: ANDOR SYSTEMS, INC. [US/US]; 10131 Bubb Road, Cupertino, CA 95014 (US).

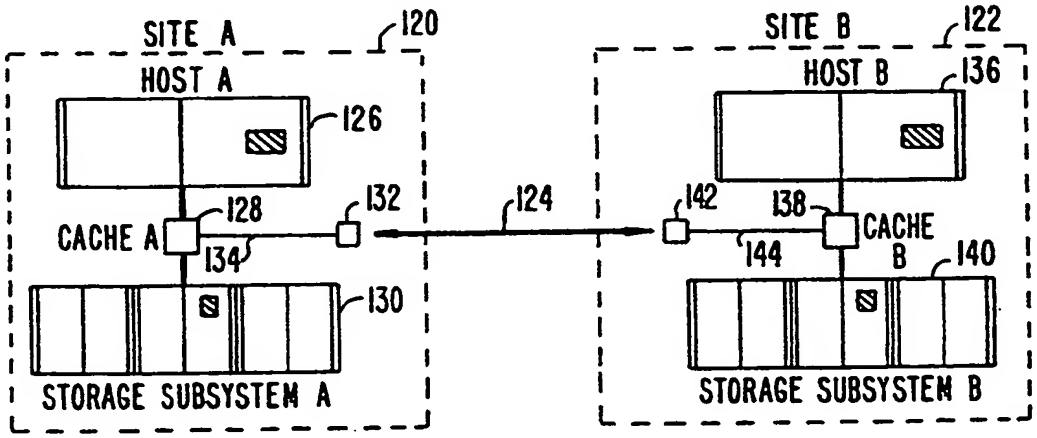
(72) Inventors: BERGSTEN, James ; 12231 Country Squire Lane, Saratoga, CA 95070 (US). KING, David ; 1700 Halford Avenue, Ap. 112, Santa Clara, CA 95051 (US). NADZAM, William ; 326 Creekwood Court, Morgan Hill, CA 95037 (US). BODWIN, James ; 22475 Palm Avenue, Cupertino, CA 95014 (US).

(74) Agents: ALBERT, Phil, H. et al.; Townsend and Townsend Khourie and Crew, Steuart Street Tower, 20th Floor, One Market Plaza, San Francisco, CA 94105 (US).

(81) Designated States: AU, CA, JP, KR, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: REMOTE DUAL COPY OF DATA IN COMPUTER SYSTEMS



(57) Abstract

A method and apparatus for achieving remote dual copy of data in a computer system (100) provides for data from a first storage device (104) to be copied onto a second storage device (106) that is located a great distance from the first storage device (104). A cache processor (128) is used at the site of the first storage device (104) to form a record of data written to the first storage device (104), by a host processor (102) at the site. The record is transmitted to a second cache processor (138) at the distant site. The second cache processor (138) uses the record to update the second storage device (106) whereby an identical copy of data on the first storage device (104) is maintained. Another aspect of the invention allows data to be migrated via a cache processor between storage devices at a same location so that servicing and non-disruptive tape backup may be performed on one of the devices.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**FR**	France	**MR**	Mauritania
AU	Australia	**GA**	Gabon	**MW**	Malawi
BB	Barbados	**GB**	United Kingdom	**NE**	Niger
BE	Belgium	**GN**	Guinea	**NL**	Netherlands
BF	Burkina Faso	**GR**	Greece	**NO**	Norway
BG	Bulgaria	**HU**	Hungary	**NZ**	New Zealand
BJ	Benin	**IE**	Ireland	**PL**	Poland
BR	Brazil	**IT**	Italy	**PT**	Portugal
BY	Belarus	**JP**	Japan	**RO**	Romania
CA	Canada	**KP**	Democratic People's Republic of Korea	**RU**	Russian Federation
CF	Central African Republic			**SD**	Sudan
CG	Congo	**KR**	Republic of Korea	**SE**	Sweden
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovak Republic
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LU**	Luxembourg	**TD**	Chad
CS	Czechoslovakia	**LV**	Latvia	**TG**	Togo
CZ	Czech Republic	**MC**	Monaco	**UA**	Ukraine
DE	Germany	**MG**	Madagascar	**US**	United States of America
DK	Denmark	**ML**	Mali	**UZ**	Uzbekistan
ES	Spain	**MN**	Mongolia	**VN**	Viet Nam
FI	Finland				

BNSDOCID: <WO_____9400816A1_I_>

REMOTE DUAL COPY OF DATA IN COMPUTER SYSTEMS

NOTICE REGARDING COPYRIGHTED MATERIAL

A portion of the disclosure of this patent document contains material which is subject to copyright protection. The copyright owner has no objection to the facsimile reproduction by anyone of the patent document or the patent disclosure as it appears in the Patent and Trademark Office patent file or records, but otherwise reserves all Copyright Rights whatsoever.

FIELD OF INVENTION

This invention relates generally to the transfer of data in a computer system and specifically to transfers of data such that copies of a data volume are maintained on separate physical storage devices.

BACKGROUND OF THE INVENTION

Modern computer systems handle extremely valuable and often irreplaceable data. Usually the data is stored in a random access memory (RAM) which allows fast access to the data but which requires a continuous power supply or the data will be lost. This requirement makes RAM a "volatile" storage medium. Data is usually stored in other, nonvolatile, storage media for more permanent storage when it is not immediately needed by the computer processor.

There are several types of nonvolatile storage media such as magnetic tape, magnetic disks, optical disks, bubble memory, etc. These types of nonvolatile storage will retain data even in the absence of electrical power and are characterized by longer access times and larger capacity than the solid state memory storage devices such as RAM. While non-volatile storage media provide for safer, more permanent storage of data, the data in nonvolatile storage media is still susceptible to loss or corruption due to failures in the media. In nonvolatile storage devices the failure is usually in the form of a physical or

mechanical failure of the apparatus which writes to, or reads from, the nonvolatile storage device.

In the case of a magnetic disk, an electrical or magnetic field may cause the magnetization of the disk surface to become damaged. Alternatively, the surface of the magnetic media can be subjected to physical force that could cause the surface magnetic material to be sheared or removed from the magnetic disk. Often these failures are the result of normal wear and tear during the life cycle of the storage device. However, they can also be caused by movement of the devices, for example, in transporting the devices, or by natural disasters such as fires and earthquakes. Other storage media are also prone to physical disruption.

Because of the possibility of a failure of the storage device the data on the storage device is often "backed up." This involves copying the data on the first storage device onto a second storage device. Thus, a volume of data on a disk drive might be backed up onto a magnetic tape or onto a second magnetic disk.

The storage device backups can be performed "online" or "offline." Online backups are performed by allowing the host computer to access the data on the first storage device while a backup or copying of the first storage device is taking place. Unfortunately, because the copy of the first storage device data is being created while the host is accessing and changing the data on the first storage device the second storage device containing the copy will most likely never contain a "snapshot" of the data on the first storage device. That is, the data on the second storage device, intended to be a copy of the data on the first storage device at some point in time, will instead be a collection of parts of the data of the first storage device which never existed simultaneously. This makes it difficult to reconstruct the data in the system at a later time.

To illustrate this problem, assume that storage device 1 has three data elements or parts A, B and C. Assume that a copy is being made of storage device 1 onto storage device 2 while at the same time allowing the host processor to access and modify the data on storage device 1 (i.e., online backup).

Copying of data part A onto storage device 2 completes and copying of data part B subsequently begins. During the copying of data part B onto storage device 2 the host processor modifies data part A on storage device 1 to be data part D on storage device 1 and modifies data part C on storage device 1 to be data part E on storage device 1. Next, data part B from storage device 1 is copied to storage device 2, after which the copying of what was formerly data part C begins. However, since data part C has been changed to data part E the backup operation instead copies data part E to storage device 2.

Storage device 2 ends up containing data parts A, B and E. However this collection of data parts never existed on storage device 1. Storage device 1 contained, successively, the collections A, B, C; D, B, C and D, B, E. The resulting "copy" of storage device 1 on storage device 2 of a 2A, B, E is a result of allowing the host to access storage device 1 during the copy or backup operation.

This inability to insure an exact copy of the state of all of the data on storage device 1 at some instant in time may make it difficult, if not impossible, to reproduce or recreate data which the host processor was manipulating.

Offline tape backups also have related problems. In an offline backup the host processor is not allowed to modify data on the storage device being copied. This insures that an exact copy of the state of data on storage device 1 at some instant in time will be obtained. However, because of the large capacity of nonvolatile storage devices, and because it is often desired to perform frequent backups, the offline backup process means that the computer system is not operable at full capacity, or perhaps not at all, for the duration of the offline backup. In some computer system applications such as airline reservations, banking, etc., such a shutdown would not be an option.

During a backup, storage device 1 and storage device 2 are in close proximity to each other. This is because there is a propagation delay in electrical signals across cables that run between first and second storage devices, through a control unit and, in the case of online backups, to a host processor.

BNSDOCID: <WO_____9400816A1_I_>

The length of cables required results in a non-negligible delay in signal transfers that affect the efficiency of data transfers from storage device to storage device, from storage device to a control unit and from a storage device to a host processor. Typically the maximum cable length is about twenty-five feet between a device that reads from or writes to a storage device and the storage device itself.

Consequently, the backup schemes described above are not useful in protecting against a "disaster event" such as a fire which would destroy the first storage device and second (copy) storage device if the two storage devices are kept in the same general area.

The second storage device can be removed to a physically remote location after the backup operation is completed. However this does not protect against a disaster event occurring during the copy operation. Further, the physical transport of storage devices or storage media is expensive and could damage the media. Therefore there arises a need for an invention that provides for the copying of data on a first storage device to a second remotely located storage device.

SUMMARY OF THE INVENTION

The present invention is directed to a method and apparatus for backing up data to physically distant media without disrupting normal processing and without risk of physical damage to the storage media. In one embodiment of the present invention a cache unit comprising a cache memory and processor are disposed in the data path between the host processor and the storage devices. Any data exchanged between the host processor and the storage devices also passes through the cache unit. The cache unit appears to the host processor as a very fast storage device. That is, the cache unit receives commands intended for a storage device and returns data requested by the host processor from selected storage devices. The storage devices, in turn, see the cache unit as a host processor since the cache unit issues read and write operations to the devices, which emulate the host processor's requests and protocol.

The cache unit is able to respond to host operations intended for storage devices much more quickly than the storage device itself. This is because the cache unit employs very fast memory so that any write or read operations performed by the cache unit is carried out in a fraction of the time. Also, since the cache unit receives all data going between the host and the storage device, the cache unit is able to create a record of operations performed on data stored on a given storage device. This record can subsequently be transmitted to a storage device in a remote location. A second cache unit at the remote location then uses the record to update a second storage device at the remote location so that a copy of the data on the first storage device is up-to-date. In this way, a second copy of the data on the first storage device is maintained at a remote location.

The cache unit also allows for so-called "non-disruptive" or "transparent" backup on data on a storage device. The cache unit accomplishes this by "migrating" data between a first storage device and a second storage device so that a copy of the first storage device's data is formed on the second storage device. The data migration is done without affecting the host processor's operations or resources. Once a copy is formed, either the first or second storage device can be used to perform the remote backup while the other storage device is used in host operations.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will now be further described in conjunction with the attached drawings as follows:

Figure 1 is a system block diagram of a particular embodiment of a computing system according to the present invention;

Figure 2 is a block diagram of a particular embodiment of an apparatus for achieving remote dual copy;

Figure 3 is a block diagram of an alternative embodiment of an apparatus according to the present invention for providing remote dual copy among four sites;

Figure 4 is a block diagram of another alternative embodiment of an apparatus according to the present invention for providing remote dual copy among four sites.

Figure 5 is a system block diagram of a second embodiment of a computing system according to the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Figure 1 shows a system configuration 100 of a host processor 102 which exchanges data with storage device A and storage device B.

Storage devices A and B are connected to a control unit 108. In prior art systems control unit 108 is connected directly to host 102. Control unit 108 handles the communications protocol between host 102 and storage devices A and B. Control unit 108 also provides some buffering of data between host 102 and storage devices 104 and 106.

In the present invention, a cache processing unit 110 (also referred to as a "cache processor") is placed in the data path between host 102 and control unit 108. Thus, the "front end" of the cache processing unit 110 is connected to host processor 102 by means of a data path 112 and the "back end" of the cache processing unit 110 is connected to the control unit 108 through data path 114.

Cache processing unit 110 may be constructed in accordance with the teachings of U.S. Patent Application No. (unknown) entitled "Intermediate Processor Disposed Between a Host Processor Channel and a Storage Director," which is herein incorporated by reference as if set forth in full in this document.

Cache processing unit 110 provides a large amount of fast buffering between host processor 102 and the storage devices A and B. The front end of the cache processor is connected to the host by data path 112 and emulates the communications protocol of control unit 108 while the back end of the cache processor emulates the transmission protocol of host processor 102. In this manner, the insertion of the cache processor between host processor 102 and control unit 108 is transparent to

both the host processor and control unit. Therefore, no additional modifications to the software or hardware in the initial system are necessary to accommodate the cache processor.

The buffering provided by cache processing unit 110 is accomplished in high-speed memory which allows cache processing unit 110 to respond to host-issued operations much more quickly than control unit 108 and storage devices A and B. In this manner, cache processing unit 110 processes host operations to the storage devices and improves the system performance.

For example, in a host write operation host 102 sends data to the cache processor to be written to one of storage devices A or B. The cache processor stores this data in a buffer and informs the host that the write operation is complete. The cache processor often indicates that the write operation is complete even before the data has actually been written to storage devices A or B.

The data stored in the cache processor's buffer is in the form of a "record". In the preferred embodiment, the record is a variable size and comprises the data to be written along with control information such as, e.g., a checksum, record identifier, count field, etc. In other embodiments it will be possible to improve the efficiency of the cache processor's performance by converting the data to be written into a record of changes to the data on a storage device. For example, the host may be writing or changing a small portion of data (a record) residing within a larger portion of data (a track) on storage device A. Often the host will do this by writing the entire track which means that parts of the data sent by the host will be redundant since old data will be overwritten by new data that is actually identical to the old data. The cache processor would detect such instances of writing redundant data and eliminate them from the record of changes.

The cache processor could also improve efficiency when the host performs two write operations to a record on a storage device without doing an intervening read operation. Since the first write operation is unnecessary (it will be overwritten by the second write operation and never used) the cache processor may simplify the write operations by discarding the first write

operation and retaining only the data from the second write operation as part of the record.

In the preferred embodiment, in the case of a read operation by the host from a storage device, the cache processor provides the data if it is resident in a cache processor buffer. If the requested data is not present in the cache processor, the cache processor accesses the storage device to obtain the data and then provide it to the host. The cache processor optimizes the data in its buffers so that data that is most likely to be used in the future will be more likely to be resident in a cache processor buffer. Cache processing unit 110 also performs other functions such as keeping copies of data in nonvolatile memory and performing error detection, correction and reporting.

Figure 2 shows a particular embodiment of an apparatus for achieving remote dual copy. Figure 2 shows a site A connected to a site B by means of a transmission line 124. Site A and site B are the physical locations of two separate computer systems that are sufficiently apart from each other so that a disaster event such as a fire or an earthquake that affects one site is unlikely to affect the other site. Transmission line 124 can be any suitable communication link such as hardwire, satellite, fiberoptic, etc.

Each of sites A and B include a computer system of the type shown in Fig. 1. More specifically, the computer system at site A comprises a host 126, a cache processor 128 and storage subsystem 130. Storage subsystem 130 contains at least one storage device such as storage device A of Figure 1. Such a storage device could be a magnetic disk drive, magnetic tape, nonvolatile memory, or other means of storage. Storage subsystem A could further comprise multiple storage devices connected to one or more control units. One such arrangement would be the arrangement between control unit 108 and storage devices 104 and 106 as illustrated in Figure 1. Any traditional arrangement of storage devices and control units are understood to be represented by the storage subsystem 130 of Figure 2 (see, for example, "IBM 3990 Planning, Installation, and Storage Administration Guide," IBM Publication number GA32-0100).

In Fig. 2 site B 122 comprises a computer system analogous to the computer system of site A. That is, site B has a host 136, cache processor 138 and storage subsystem 140. Cache processor 138 is also connected to network link 142 through datapath 144. Thus, in Fig. 2, cache processor A can send and receive data to and from cache processor B by means of the intercache link established by datapath 134, link 132, transmission line 124, link 142 and datapath 144. Thus, in Fig. 2, the capability of maintaining an identical copy at site B of data at site A (i.e., remote dual copy) is achieved by transmitting records of changes to the data at storage subsystem A by host A to cache processor B. Eventually, the record received at cache processor B is used to modify the copy of data in storage subsystem B to bring that copy up to date with the storage subsystem A data.

For purposes of illustration, we will assume the simplest case where the storage subsystems A and B each are a single storage device. Further, we assume that the storage device is a magnetic disk drive. However, the present invention is understood to apply to multiple storage devices in various configurations, including other than magnetic media, which comprise a storage subsystem as is commonly known in the art.

The computer system at site A could be running any application such as scientific, financial, educational, etc. The application of the system at either site A or site B is not limited in any way by the addition of the cache processors and the intercache link.

As host A writes data to storage subsystem A, cache processor A forms a record of the changes to the data in storage subsystem A. Cache processor A breaks up this record into "packets". As a packet is created, it is assigned a sequence number and sent to the network link A. Network link A transmits the packet over transmission line 124 to network link B. Network link B then transfers the packet to cache processor B over datapath 144. In cache processor B the information in the transmitted packet is used to rebuild the record identically to the record in cache processor A. Subsequently, this record is used to update the data in storage subsystem B. Note that

a packet can be a very small portion of a record such as one or more characters or bytes of data. Alternatively, the packet can be rather large containing hundreds, thousands or even millions of bytes. Typically, the packet size depends on the characteristics of the intercache link. Therefore, if the error rate over the link is high because the distance is long or because the link is accomplished by electromagnetic transmissions that are susceptible to noise, the packet size will be small. The details of implementing a packet-switched network are well known in the art. Further, any type of communications protocol can be used to establish data transfer over the intercache link from cache processor A to cache processor B as shown in Fig. 2.

The sequence number associated with a packet when it is created at cache processor A is used to form the completed record from several packets and to detect and correct errors in transmission of packets between cache processor A and cache processor B. In the preferred embodiment, the sequence numbers are incremented successively and the network link B will keep track of packets as the packets arrive. Thus, if packets with sequence numbers 1, 2 and 4, respectively, are received by network link B, network link B will retain the packets until a packet with sequence number 3 is received, which will complete the ordered transmission of a record. Network link B, therefore, will not send an incomplete record to cache processor B but instead will wait until all packets for a record have been received.

It is possible for cache processor B to receive packets that are not in order and to assemble records from the packets. However, assembling the record within the network link B results in a more efficient communications system design. Any manner of error detection and correction, communications protocol and physical transmission line may be used to implement the details of the intercache link design.

If the intercache link has a high bandwidth, e.g., as with an optical fiber link, the rate of data transfer from cache processor A to cache processor B can be quite large. This allows for frequent transmissions of records from cache processor A to cache processor B. This means that the copy at site B of data at

site A will not "lag" too far behind the updates to the data of storage subsystem A.

In the case described above where storage subsystem A is a single magnetic hard disk and where storage subsystem B is a second magnetic hard disk and is intended to be a copy of the disk of storage subsystem A, all of the write operations from host A to the storage subsystem A disk will be sent from cache processor A to cache processor B so that the disk at storage subsystem B can be updated identically with the disk at storage subsystem A. However, another aspect of the present invention allows a storage device to be "partitioned" into two or more areas.

When partitioning is implemented, storage space on the storage device may be allocated into multiple partitions, and a subset of one or more of the total partitions can be designated to be backed up or copied to site B. The non-designated partitions may be modified by host A without transmitting records of the changes to site B. Also, if only a portion of the magnetic disk at site B is needed to maintain a backup copy of the disk at site A, then the host B processor can write to those portions of the magnetic disk at storage subsystem B that are not needed to maintain the copy of data from subsystem A. In other words, when the amount of data being remotely copied does not require the full resources of the storage media of a site, then the host processor may access the remaining media without the cache processor having to form and transmit records of changes. The specific packet size, transmission speed, frequency of packet transmission, etc. will vary with the specific nature of the transmission line, equipment and protocol being utilized.

A "map" of the storage media showing which portions of storage are dedicated to remote dual copy is maintained by the cache processor at the site of the respective storage media. The maintenance of this map in the cache processor allows the host and storage subsystem at the transmitting site, site A 120 in Fig. 2, to operate without regard to the remote dual copy, which takes place transparently to the system at site A. At site B the host processor 136 can operate independently, that is, without regard to the remote dual copy insofar as its storage resources

BNSDOCID: <WO_____9400816A1_I_>

are not allocated to maintaining a remote copy of data from site A. This is possible since cache processor B maintains a map of storage subsystem B's storage media allocation. Cache processor B will make storage subsystem B's resources available to host B only insofar as storage subsystem B's resources are not being used to perform the remote dual copy.

In Fig. 2, site A 120 is the "primary" site and site B 122 is the "secondary" site. A primary site is that site whose cache processor transmits a record of write operations or changes to a receiving or secondary site. The primary site may update its storage subsystem media before the secondary site, although this need not always be the case. Additionally, the host processor at the primary site has full use of its storage subsystem resources while the host processor at the secondary site B may only use its storage subsystem resources insofar as they are not being used to back up the primary site's storage subsystem.

The primary and secondary roles may be reversed between sites A and B of Fig. 2. In that case changes are sent from cache processor B to cache processor A over the transmission line. Thus, identically with the foregoing discussion, host B would write changes to cache processor B that would be transmitted to cache processor A. The switching of primary and secondary roles, assuming that network link A, transmission line 124, and network link B implement a "two-way" or bi-directional transmission link, is accomplished by setting parameters within both cache processor A and cache processor B.

The switching of primary and secondary roles could be initiated by an instruction from host processor A or by other means, such as manually by an operator at one of the sites, or automatically by one of the caches themselves. This switching of roles would be useful where the demands on the computer systems of sites A and B are such that site A is operating at high capacity while site B is at low capacity and, later, where site B is at high capacity when site A is at low capacity. In this situation, site A would be designated the primary when operating at high capacity and site B would later be designated primary when operating at its high capacity. This provides a

way to maintain remote dual copy of data at each site while at the same time minimizing the interference with resources at each site.

Fig. 3 shows an embodiment of the present invention as implemented in a four-site system. In Fig. 3 site A 140 and site B 142 are identical with site A 120 and site B 122 of Fig. 2 with the exception that they are not connected by an intercache link. Site C 143 of Fig. 3 also has equivalent components to those discussed for sites A and B in regard to a discussion for Fig. 2. Site D 144 of Fig. 3 has three network links 150, 152 and 154 connected to cache processor 156. Site D also shows storage devices 158, 160 and 162 connected to storage subsystem 164 of site D 144. Storage devices 158, 160 and 162 are shown as separate blocks in Fig. 3 for ease of discussion. However, these storage devices are conceptually part of storage subsystem 164 since, as described above, the storage subsystem symbol used to represent, for example, storage subsystem 164, is intended to represent any configuration of storage devices, control units, storage directors, etc. Also in Fig. 3 storage devices 158, 160 and 162 are shown connected to storage subsystem 164 in a "daisy chain" fashion. This, also, is merely a representative illustration and any manner of connection among storage devices 158, 160 and 162 in storage subsystem 164 is considered within the scope of the invention, e.g., "ring", "multipoint", "Star Network", etc.

In Fig. 3 site D 144 acts as the repository for maintaining remote copies of data on storage devices at each of the other sites A, B and C. Thus, data written by host processor 166 of site A to storage subsystem 168 is first received by cache processor 170. Cache processor 170 later transmits a record of the changes or writes through the intercache link between cache processors 170 and 156. This link comprises a datapath 171, a network link 173, a transmission line 175, a second network link 150, and a second datapath 151 as shown in the illustration.

The record received from cache processor 170 by cache processor 156 is used to update a copy of site A's storage device (not shown, but residing in storage subsystem 168) on storage device 158 which is designated as a remote copy device for site

A. Similarly, site B 142 will send data that is used by cache processor 156 to update a copy of site B's data on storage device 160. Site C's data is backed up on storage device 162.

As Fig. 3 illustrates, several sites can use a common site as a repository for purposes of remote dual copy.

Fig. 4 illustrates a more complex arrangement of dual remote copy among four sites. In Fig. 4, data from site A 170 is copied and maintained by the present invention in the manner described above. Site A's data is transmitted and used to maintain copies on storage device A 172 of site B 174 and on storage device 176 of site D 178. Fig. 4 also shows site B 174 transmitting data to site A 170 and site C 180. Site A maintains a copy of site B's data on storage device 182 while site C maintains a copy of site B's data on storage device 184. As further shown in Fig. 4, site A maintains a copy of site D's data on storage device 186. As the reader can easily determine from Fig. 4, site A maintains copies of sites B and D, site B maintains copies of sites A and C, site C maintains copies of sites B and D, and site D maintains copies of sites A and C.

Note that in Fig. 4 each of the network links such as 200 and 202 will both transmit and receive data. Also, cache processor 204 at site A and cache processor 206 at site B must each act, at different times, as primary and secondary sites. For example, in order for site B to maintain a copy of site A's data on storage device 172, the cache processor 204 of site A 170 must transmit a record of site A's data through network link 200 over transmission line 207 to network link 202. This record is then received by cache processor 206 which ultimately updates storage device 172. Also, in order for site A to maintain a copy of site B's data on storage device 182, cache processor 206 must transmit a record through network link 202, transmission line 207, and network link 200, to finally be received at cache processor 204 which ultimately updates storage device 182. Thus, it is seen that cache processor 204 operates as a primary when cache processor 206 is a secondary with respect to data transfers between sites A and B over transmission line 207; and cache processor 206 acts as a primary while cache processor 204 acts as

BNSDOCID: <WO_____9400816A1_I_>

a secondary at other times during data transfers between sites B and A.

Fig. 4 illustrates that many configurations of sites and primary and secondary roles are possible. As an alternative to direct data transmission from site A to site B over transmission line 207, transmission could be via lines 208, 210 and 212 which would mean that a data transmission from A to B would also pass through the cache processors at sites D and C. The present invention is understood to include all such site configurations and primary/secondary relationships as can be supported by communications network technology. The illustrations herein are only but a few examples of possible site configurations for implementing the remote dual copy feature of the present invention.

The present invention also provides a way to perform "migration" of data between storage devices at a single site. This allows a copy of a storage device to be made at the site so that one of the devices, either the one containing the original data or the copy of data, may be backed up or serviced.

Fig. 5 illustrates the data migration feature of the present invention. Fig. 5 shows a host 302 connected to a cache processor 304. Cache processor 304 is connected to two control units, control unit 306 and control unit 308. Each control unit has two storage devices connected to it. Storage device A 310 and storage device B 312 are connected to control unit 306 while storage device C 314 and storage device D 316 are connected to control unit 308.

Assuming that storage device A is desired to be backed up or serviced, a command would be given to cache processor 304 to begin data migration. The command could be issued by the host processor or by an operator. The command would designate a storage device not being used to be the target device for the data migration. For example, device D could be an empty disk recently mounted on a disk drive to be the target of the migration.

The cache processor migrates data by performing read operations from device A and transferring the data read to device D via write operations. This migration occurs without

interrupting the host operations since the cache processor performs the migration operations when host operations do not require the use of datapaths 318 and 320.

During migration the host may be changing data on device A. The cache processor keeps a copy of these changes and updates data that has been written to device D if the data written to device D is a copy of data on device A that has since been modified. By updating data on device D as it is changed on device A, the cache processor will eventually create an exact copy of device A's data on device D.

After a copy of device A has been formed on device D, device A is "frozen" by the cache processor. That is, all attempts by the host to access data at device A are routed to device D. This allows the host to have uninterrupted access to the data as if device A were still present. As the host makes changes to the data on device D, the cache processor keeps a record made.

Once device A has been frozen, it may now be removed from the system by disconnecting datapath 322. Thus, device A can be serviced or backed up to tape offline without compromising the performance of the computer system shown in Fig. 5. In this manner, non-disruptive tape backup is achieved.

In the foregoing specification, the invention has been described with reference to a specific exemplary embodiment thereof. It will, however, be evident that various modifications and changes may be made thereunto without departing from the broader spirit and scope of the invention as set forth in the appended claims. The specification and drawings are, accordingly, to be regarded in an illustrative rather than a restrictive sense and the invention is not intended to be limited except as indicated in the appended claims.

BNSDOCID: <WO_____9400816A1_I_>

WHAT IS CLAIMED IS:

1. A method for changing first and second identical copies of data on respective first and second storage devices comprising the steps of:

communicating first change data for changing the first and second copies of data from a first host to a first cache processor unit;

receiving the first change data at the first cache processor unit;

transmitting the first change data from the first cache processor unit to a second cache processor unit over a first transmission line;

receiving the first change data at the second cache processor unit through the first transmission line;

changing the first copy of data with the first change data received at the first cache processor unit; and

changing the second copy of data with the first change data received at the second cache processor unit so that the first and second copies of data remain identical.

2. The method according to claim 1 further comprising the steps of:

communicating second change data from a second host to the second cache processor unit;

receiving the second change data at the second cache processor unit;

transmitting the second change data from the second cache processor unit to the first cache processor unit; and

receiving the second change data at the first cache processor unit.

3. The method according to claim 2 further comprising the step of changing the first copy of data with the second change data received at the first cache processor unit.

4. The method according to claim 3 further comprising the step of changing the second copy of data with the second

change data received at the second cache processor unit so that the first and second copies of data are identical.

5. The method according to claim 4 wherein the step of transmitting the second change data comprises the step of transmitting the second change data from the second cache processor unit to the first cache processor unit through the first transmission line.

6. The method according to claim 4 further comprising the steps of:

inhibiting the transmission of first change data to the second cache processor unit while the second change data is being transmitted to the first cache processor unit; and

inhibiting the transmission of second change data to the first cache processor unit while the first change data is being transmitted to the second cache processor unit.

7. The method according to step 1 further comprising the steps of:

transmitting the first change data from the first cache processor unit to a third cache processor unit; and

receiving the first change data at the third cache processor unit.

8. The method according to claim 7 further comprising the step of changing a third copy of data with the first change data received at the third cache processor unit so that the first and third copies of data are identical.

9. The method according to claim 8 wherein the step of transmitting the first change data to the third cache processor unit comprises the step of transmitting the first change data to the third cache processor unit through a second transmission line.

10. The method according to claim 9 wherein the step of transmitting the first change data to the third cache

processor unit comprises the step of transmitting the first change data to the third cache processor unit directly through the second transmission line.

11. The method according to claim 9 wherein the step of transmitting the first change data to the third cache processor unit comprises the steps of:

transmitting the first change data to the second cache processor unit through the first transmission line; and

transmitting the first change data to the third cache processor unit from the second cache processor unit to the third cache processor unit through the second transmission line.

12. The method according to claim 1 further comprising the steps of:

communicating second change data from a second host to a third cache processor unit;

receiving the second change data at the third cache processor unit;

transmitting the second change data from the third cache processor unit to the second cache processor unit; and

receiving the second change data at the second cache processor unit.

13. The method according to claim 12 further comprising the step of changing a third copy of data with the second change data received at the second cache processor unit.

14. The method according to claim 13 further comprising the step of changing a fourth copy of data with the second change data received at the third cache processor unit so that the third and fourth copies of data are identical.

15. A method of disconnecting a first storage unit from a host processor which performs read/write operations on a first copy of data stored on the first storage unit comprising the steps of:

routing read/write operations for the first copy of data from the host processor through a cache processor unit coupled to the first storage unit, the cache processor unit including a cache processor memory;

reading data from the first storage unit into the cache processor memory;

writing the data read from the first storage unit from the cache processor memory into a second storage unit coupled to the cache processor memory for creating a second copy of the first copy of data;

disconnecting the first storage unit from the cache processor unit; and

performing read/write operations from the host processor for the first copy of data on the second copy of data after disconnecting the first storage unit.

16. The method according to claim 15 further comprising the steps of:

receiving a write operation for the first copy of data before the first storage unit is disconnected from the cache processor unit; and

performing the write operation on the first and second copies of data.

17. The method according to claim 16 further comprising the steps of:

storing a record of write operations received from the host processor in the cache processor memory after disconnecting the first storage unit;

reconnecting the first storage unit to the cache processor unit; and

performing the write operations stored in the record on the first copy of data after reconnecting the first storage unit.

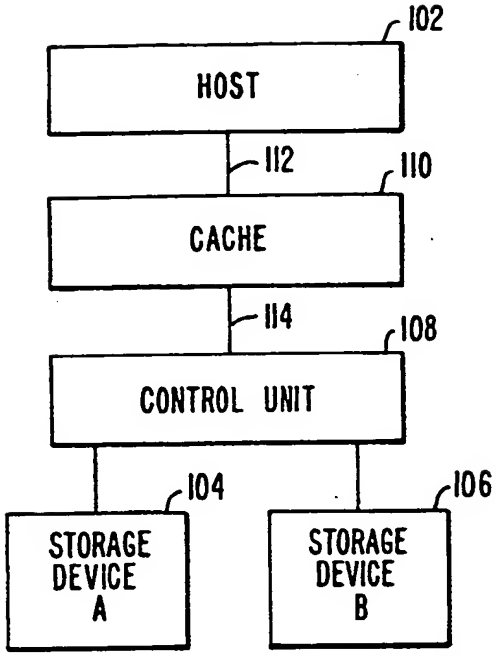


FIG. 1.

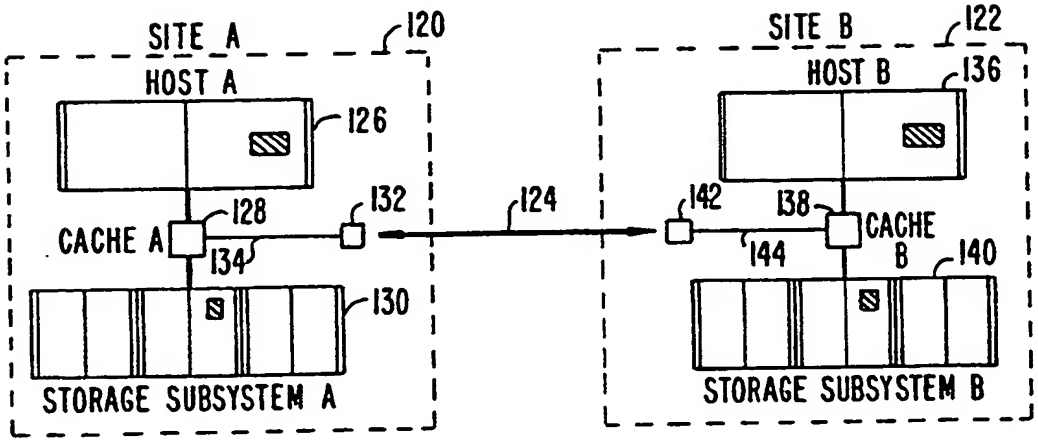


FIG. 2.

BNSDOCID: <WO_____9400816A1_I_>

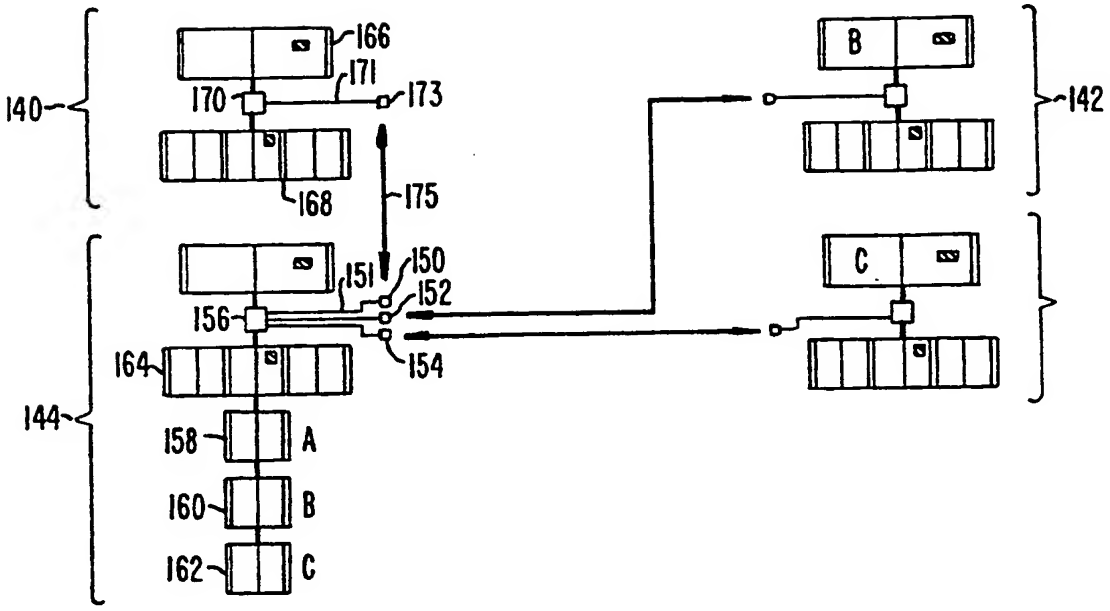


FIG. 3.

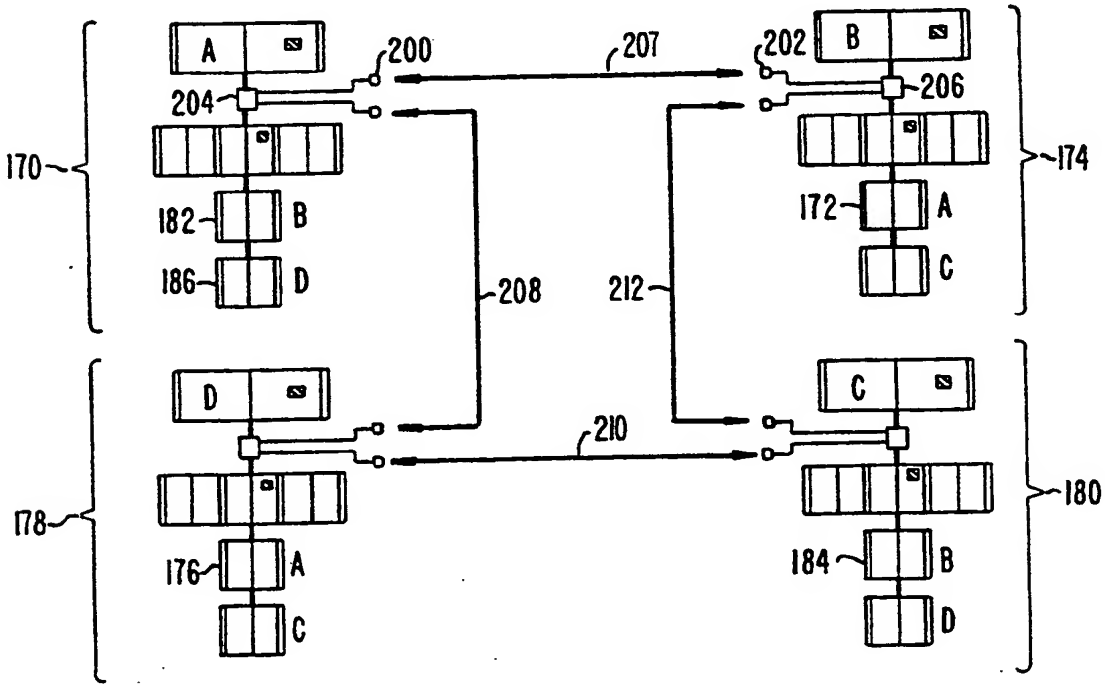


FIG. 4.

SUBSTITUTE SHEET

BNSDOCID: <WO_____9400816A1_I_>

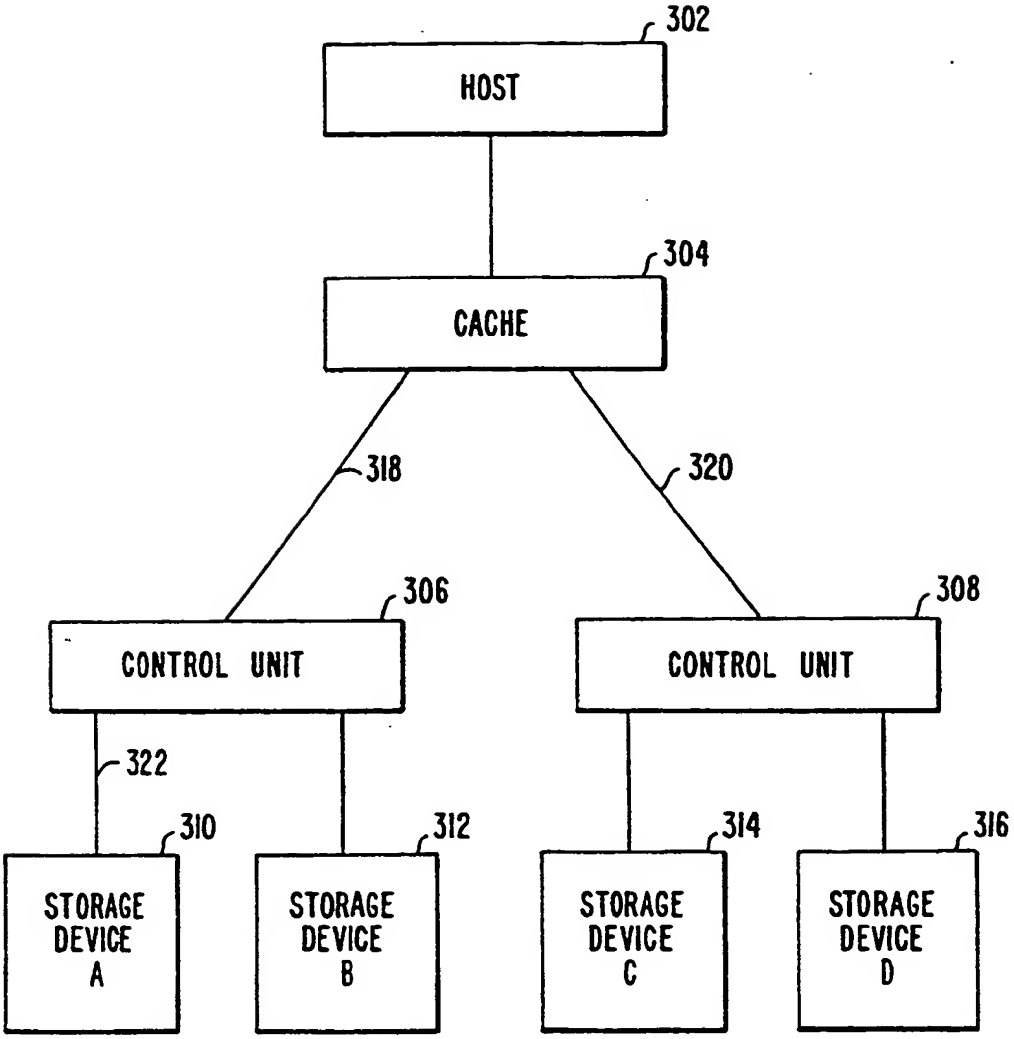


FIG. 5.

BNSDOCID: <WO_____9400816A1_I_>

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US93/05853

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) :G06F 13/00; 12/00, 12/08, 12/16, 11/00
US CL : 395/425, 575; 371/7,9,1,10.1
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 395/425, 575; 371/7,9,1,10.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US,A 4,045,781(LEVY ET AL.) 30 AUGUST 1977 NOTE ABSTRACT, COL 13 LINES 20-60.	1-17
A	US,A 4,819,159(SHIPLEY ET AL) 4 APRIL 1989 NOTE ABSTRACT	1-17
E	US,A 5,197,148 (BLOUNT ET AL.) 23 MARCH 1993 NOTE ABSTRACT	1-17
E	US,A 5,222,217 (BLOUNT ET AL.) 22 JUNE 1993 NOTE ABSTRACT.	1-17
E	US,A 5,155,845 (BEAL ET AL.) 13 OCTOBER 1992 NOTE ABSTRACT	1-17

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be part of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 30 August 1993	Date of mailing of the international search report 07 OCT 1993
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. NOT APPLICABLE	Authorized officer ALYSSA BOWLER Telephone No. (703) 305-3818

Form PCT/ISA/210 (second sheet)(July 1992)*

BNSDOCID: <WO_____9400816A1_I_>

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US93/05853

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

APS TEXT SEARCH (COPY OR (BACK-UP)
REMOTE REDUNDANT?
DISTANT DISASTER
DISK COMPUTER
CACHE

BNSDOCID: <WO_____9400816A1_I_>

THIS PAGE BLANK (USPTO)

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☑ FADED TEXT OR DRAWING
- ☑ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☑ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

THIS PAGE BLANK (USPTO)